ON THE REGULARITY OF THE CONDITIONAL DISTRIBUTION OF THE SAMPLE MEAN

VICTOR CHULAEVSKY

Abstract. We show that the hypothesis of regularity of the conditional distribution of the empiric average of a finite sample of IID random variables, given all the sample "fluctuations", which appeared in our earlier manuscript [1] in the context of the eigenvalue concentration analysis for multi-particle random operators, is satisfied for the uniform probability distributions. It extends the well-known property of Gausssian IID samples.

1. Introduction

In a few talks given at workshops on disordered quantum systems, I have mentioned a simple result of the elementary probability theory which has an interesting application to the multi-particle Anderson localization theory. It is difficult to say if the result itself is original; personally, I would be glad to learn that it is not, and to find some bibliographical reference, for it is indeed hard to believe that the elementary probabilistic problem in question never addressed, for example, in statistics. However, I am unaware of any such published (or folkloric) result.

The goal of this short note is to fill this gap and provide an elementary proof of the regularity (with high probability) of the conditional sample mean of a finite sample of uniformly distributed IID random variables, given the sigma-algebra of "fluctuations".

2. Prelude: Gaussian IID samples

Consider a sample of N IID random variables with Gaussian distribution $\mathcal{N}(0,1)$, and introduce the sample mean $\xi = \xi_N$ and the "fluctuations" η_i around the mean:

$$\xi_N = \frac{1}{N}\sum_{i=1}^{N} X_i, \quad \eta_i = X_i - \xi_N, \;\; i = 1, \ldots, N.$$

It is well-known from elementary courses of the probability theory that ξ_N is independent from the sigma-algebra $\mathfrak{F}_\eta$ generated by $\{\eta_1, \ldots, \eta_n\}$ (the latter are linearly dependent, and have rank $N-1$). To see this, it suffices to note that η_i are all orthogonal to ξ_N with respect to the standard scalar product in the linear space formed by $X_1, \ldots, X_N$ given by

$$\langle Y, Z\rangle := \mathbb{E}\,[\,Y\,Z\,]$$

where Y and Z are (real) linear combinations of $X_1, \ldots, X_N$ (recall: $\mathbb{E}\,[\,X_i\,] = 0$).

Therefore, the conditional probability distribution of ξ_N given $\mathfrak{F}_\eta$ coincides with the unconditional one, so $\xi_N \sim \mathcal{N}(0, N^{-1})$, thus ξ_N has bounded density

$$p_\xi(t) = \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi N^{-1}}} \le \frac{N^{1/2}}{\sqrt{2\pi}}.$$

Moreover, for any interval $I \subset \mathbb{R}$ of length $|I|$, we have

$$\operatorname{ess\,sup} \mathbb{P}\left\{ \xi_N(\omega) \in I \,\middle|\, \mathfrak{F} \right\} = \mathbb{P}\left\{ \xi_N(\omega) \in I \right\} \le \frac{N^{1/2}}{\sqrt{2\pi}}\,|I|.$$

The essential supremum in the above LHS is a bureaucratic tribute to the formal rule saying that $\mathbb{P}\{\,\cdot\,|\,\mathfrak{F}\,\}$ is a random variable (which is $\mathfrak{F}$-measurable), and as such is defined, generally speaking, only up to subsets of measure zero.

In this particular case – for Gaussian samples – the conditional regularity of the sample mean ξ_N given the fluctuations $\mathfrak{F}$ is granted, but is not always so, as shows the following elementary example where the common probability distribution of the sample X_1, X_2 is just excellent: $X_i \sim \mathrm{Unif}([0,1])$, so X_i admit a compactly supported probability density bounded by 1. Indeed, in this simple example, set

$$\xi = \xi_2 = \frac{X_1 + X_2}{2}, \;\; \eta = \eta_1 = \frac{X_1 - X_2}{2}.$$

The random vector (X_1, X_2) is uniformly distributed in the unit square $[0,1]^2$, and the condition $\eta = c$ selects a straight line in the two-dimensional plane with coordinates (X_1, X_2), parallel to the main diagonal $\{X_1 = X_2\}$. The conditional distribution of ξ given $\{\eta = c\}$ is the uniform distribution on the segment

$$J_c := \{(x_1, x_2) : \; x_1 - x_2 = 2c, \, 0 \le x_1, x_2 \le 1\}$$

of length vanishing at $2c = \pm 1$. For $|2c| = 1$, the conditional distribution of ξ on J_c is concentrated on a single point, which is the ultimate form of singularity.

Yet, the good news in this example is that the conditions of singularity are quite explicit, and it is simple to assess the probability of the event that the conditional probability density of ξ given $\mathfrak{F}$ is bigger than a given threshold. In the next Section, we exploit this elementary observation in a more general case of $N \ge 2$ IID random variables uniformly distributed in $[0,1]$. The applications of the main result of Section 3 are discussed in Section 4.

3. Uniform marginal distribution

Consider a sample of N IID random variables with uniform distribution $\mathrm{Unif}([0,1])$, and introduce again the sample mean $\xi = \xi_N$ and the "fluctuations" η_i around the mean:

$$\xi_N = \frac{1}{N}\sum_{i=1}^N X_i, \quad \eta_i = X_i - \xi_N.$$

Further, consider the Euclidean space $\sim \mathbb{R}^N$ of real linear combinations of the random variables X_i with the scalar product $\langle Y, Z\rangle = \mathbb{E}\,[\,YZ\,]$. Clearly, the variables $\eta_i : \mathbb{R}^N \to \mathbb{R}$ are invariant under the group of translations

$$(X_1, \dots, X_N) \mapsto (X_1 + t, \dots, X_N + t), \;\; t \in \mathbb{R},$$

and so are their differences $\eta_i - \eta_j \equiv X_i - X_j$, $1 \le i < j \le N$. Introduce the variables

$$Y_i = \eta_i - \eta_N, \;\; 1 \le i \le N-1,$$

Then the space $\mathbb{R}^N$ is fibered into a union of affine lines of the form

$$\begin{aligned}\widetilde{\mathcal{X}}(Y) &:= \{X \in \mathbb{R}^N : \eta_i - \eta_N = Y_i,\ i \le N-1\} \\ &:= \{X \in \mathbb{R}^N : X_i - X_N = Y_i,\ i \le N-1\},\end{aligned}$$

labeled by the elements $Y = (Y_1, \dots, Y_{N-1})$ of the $(N-1)$-dimensional real vector space $\mathbb{Y}^{N-1} \cong \mathbb{R}^{N-1}$. Set

$$\mathcal{X}(Y) = \widetilde{\mathcal{X}}(Y) \cap \mathbf{C}_1 = \{X \in \mathbf{C}_1 : X_i - X_N = Y_i,\ i \le N-1\}$$

and endow each nonempty interval $\mathcal{X}(Y) \subset \mathbb{R}^N$ with the natural structure of a probability space inherited from $\mathbb{R}^N$:

- if $|\mathcal{X}(Y)| = 0$ (an interval reduced to a single point), then we introduce the trivial sigma-algebra and trivial counting measure;
- if $|\mathcal{X}(Y)| = r > 0$, then we use the inherited structure of an interval of a one-dimensional affine line and the normalized measure with constant density r^{-1} with respect to the inherited Lebesgue measure on $\mathcal{X}(Y)$.

In fact, the restriction of the sample mean $\xi_N|_{\mathcal{X}(Y)}$ provides a natural coordinate along the interval $\mathcal{X}(Y)$. Moreover, $\mathcal{X}(Y)$ with the described structure of a probability space provides a natural and convenient representation for the conditional probability distribution of the random variable ξ_N, given the sigma-algebra $\mathfrak{F}_\eta$ generated by all random variables $\{\eta_i, 1 \le i \le N\}$. If $|\mathcal{X}(Y)| > 0$, then for any $\epsilon > 0$

$$\nu_\xi(\epsilon) = \nu_\xi^{(N)}(\epsilon; \omega) := \operatorname{ess\,sup} \sup_{t \in \mathbb{R}} \mathbb{P}\left\{ \xi_N \in [t, t+\epsilon] \,\middle|\, \mathfrak{F}_\eta \right\} \le \frac{\epsilon}{|\mathcal{X}(Y)|}. \tag{3.1}$$

Theorem 1.

$$\mathbb{P}\left\{ \nu_\xi^{(N)}(s) \ge 2Ns^{1/2} \right\} \le s^{1/2}.$$

Proof. Recall that the elements $Y \in \mathbb{Y}^{N-1}$ can be considered as functions on $\mathbb{R}^N = \{X = (X_1, \dots, X_N)\}$; we will consider their restrictions to the unit cube $\mathbf{C}_1$. Introduce the random variable

$$r : X = (X_1, \dots, X_N) \mapsto |\mathcal{X}(Y(X))| \in [0, 1],$$

which is $\mathfrak{F}_\eta$-measurable and takes constant value $|\mathcal{X}(Y)|$ on each element $\mathcal{X}(Y)$.

Owing to Eqn (3.1), for any $\delta > 0$,

$$\mathbb{P}\{\nu_\xi(s) \ge \delta\} \le \mathbb{P}\left\{ \frac{s}{|\mathcal{X}(Y)|} \ge \delta \right\} = \mathbb{P}\left\{ |\mathcal{X}(Y)| \le s\delta^{-1} \right\}. \tag{3.2}$$

Let

$$X_* = X_*(X) = \min_i X_i,\ X^* = X^*(X) = \max_i X_i.$$

While $X^*(X)$ and $X_*(X)$ vary along the elements $\mathcal{X}(Y)$, their difference $X^*(X) - X_*(X)$ does not; it is uniquely determined by $\mathcal{X}(Y)$, so it makes sense to write $X^*(Y) - X_*(Y)$, meaning $X^*(Y(X)) - X_*(Y(X))$. Furthermore,

$$r(Y) = |\mathcal{X}(Y)| = 1 - (X^*(Y) - X_*(Y)) \in [0, 1],$$

so

$$\begin{aligned}\mathbb{P}\{|\mathcal{X}(Y)| \le s\} &= \mathbb{P}\{1 + X_* - X^* \le s\} = \mathbb{P}\{X_* \le s - (1 - X^*)\} \\ &\le \mathbb{P}\{X_* \le s\}.\end{aligned} \tag{3.3}$$

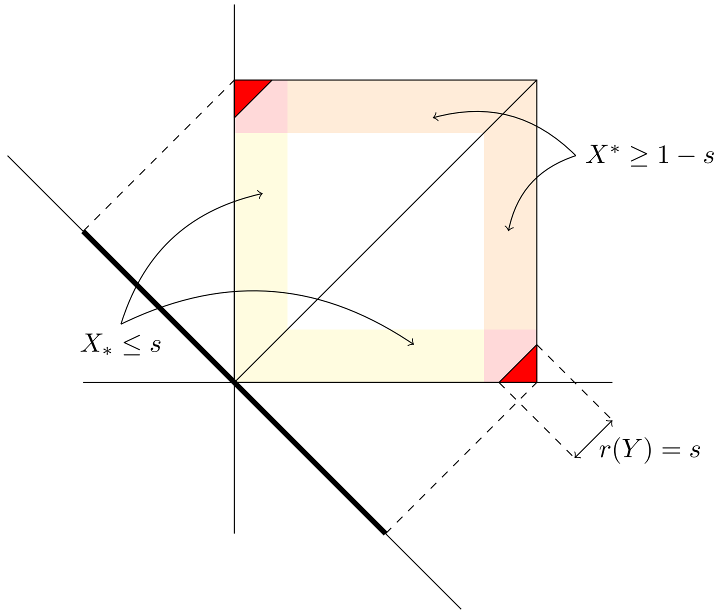

FIGURE 1. Example for Theorem 1, with $N = 2$. Yellow color corresponds to the area where $X_* \le s$, orange to $X^* \ge 1-s$, pink to their intersection, and red to $1 - X^* + X_* = r \le s$.

The latter inequality gives a crude but simple upper bound, sufficient for our purposes. In the particular case where $N = 2$, it is readily seen (cf. Fig. 1) that the exact value is $s^2/2 = O(s^2)$, while the above estimate gives only $O(s)$, $0 \le s \ll 1$.

The probability distribution of the minimum $X_* = X_*^{(N)}$ is known explicitly:

$$\mathbb{P}\left\{ \min_{1\le i\le N} X_i \ge t \right\} = \mathbb{P}\left\{ \forall\, i \in [1,N]\ X_i \ge t \right\} = \mathrm{e}^{N\ln(1-t)}.$$

Since $-2t \le \ln(1-t) \le 0$ for $t \in [0, 1/2]$, and $\mathrm{e}^{-u} \ge 1-u$ for $u \in [0,1]$, we have, for $0 \le t \le \frac{1}{2N}$,

$$\begin{aligned}\mathbb{P}\left\{ \min_{1\le i\le N} X_i \le t \right\} &= 1 - \mathrm{e}^{N\ln(1-t)} \\ &\le 1 - (1-2Nt) = 2Nt.\end{aligned} \tag{3.4}$$

Therefore, combining (3.3) and (3.4), we get

$$\mathbb{P}\left\{ |\mathcal{X}(Y)| \le t \right\} \le 2Nt. \tag{3.5}$$

Setting $\delta = 2Ns^{1/2}$ in Eqn (3.2),

$$\mathbb{P}\left\{ \nu_\xi(s) \ge 2Ns^{1/2} \right\} \le \mathbb{P}\left\{ \frac{s}{|\mathcal{X}(Y)|} \ge 2Ns^{1/2} \right\} = \mathbb{P}\left\{ |\mathcal{X}(Y)| \le \frac{s^{1/2}}{2N} \right\}$$

and using (3.5) with $t = \frac{s^{1/2}}{2N}$, we obtain

$$\le 2N \cdot \frac{s^{1/2}}{2N} = s^{1/2}. \tag{3.6}$$

Note that, in terms of Eqn (3.4), the condition $t \le 1/(2N)$ is trivially fulfilled, for in (3.6) we have $t = s^{1/2}/(2N)$, $s \le 1$. This completes the proof. $\square$

In ref. [1] was introduced the following more general condition, which actually does not assume the independence of the random variables. Let $\mathbf{C}_R(x)$ be the cube in $\mathbb{Z}^d$ centered at x and of sidelength $2R$. Consider a sample of IID random variables $V(y;\omega)$, $x \in \mathbf{C}_R(x)$, introduce as before the sample mean $\xi = \xi_{\mathbf{C}_R(x)}$ and the conditional continuity modulus $\nu_{\xi_{\mathbf{C}_R(x)}}(s)$ given the sigma-algebra of fluctuations. Then the hypothesis has the following form: for some $C', C'', A', A'', B', B'' \in (0, +\infty)$

$$\mathbb{P}\left\{ \nu_{\xi_{\mathbf{C}_R(x)}}(s) \ge C' R^{A'} s^{B'} \right\} \le C'' R^{A''} s^{B''}. \tag{3.7}$$

We see that, for an IID sample with distribution $\mathrm{Unif}([0,1])$, it is fulfilled with

$$C' = C(d),\ A' = d,\ B' = \frac{1}{2},\ \text{and } C'' = 1, A'' = 0, B'' = \frac{1}{2}.$$

Observe that our estimates are expressed in terms of the cardinality N of the sample $X_1, \dots, X_N$. In the applications to the eigenvalue concentration analysis of random operators (cf. [1, 2]) on locally finite countable graphs $\mathcal{Z}$, N has the meaning of the cardinality of a ball $\mathrm{B}_L(x) = \{y \in \mathcal{Z} :\ \mathrm{dist}_{\mathcal{Z}}(x,y) \le L\}$ of some radius L. In particular, for the balls $\mathcal{Z}$ in the class $\mathfrak{Z}(D)$ featuring a power-law growth, viz.

$$\forall\, L \ge 1 \quad \sup_{x \in \mathcal{Z}} \mathrm{card}\ \mathrm{B}_L(x) \le C_D L^D,\ \ D \in (0, +\infty),$$

for some $C_D < \infty$, we obtain the following estimate:

$$\mathbb{P}\left\{ \nu_{\xi_{\mathbf{C}_R(x)}}(s) \ge 2C_D R^D s^{1/2} \right\} \le s^{1/2}. \tag{3.8}$$

The adaptation to the uniform distribution $\mathrm{Unif}([a,b])$, $a < b$, is straightforward.

4. Applications

Consider a Hermitian random matrix $H(\omega) \in \mathrm{Mat}(N, \mathbb{C})$ of the form

$$H(\omega) = H_0 + \mathrm{diag}(V_1(\omega), \dots, V_N(\omega))$$

where $V_i(\omega)$ are IID with uniform distribution in $[0,1]$. Re-write $H(\omega)$ as follows:

$$H(\omega) = K(\omega) + \xi(\omega)\,\mathbf{1}$$

where $\xi = \xi_N$ is the sample mean of $(V_1, \dots, V_N)$ and the random operator

$$K(\omega) = H_0 + \mathrm{diag}(\eta_1(\omega, \dots, \eta_N(\omega))$$

is measurable with respect to the sigma-algebra $\mathfrak{F}$ generated by the fluctuations $\eta_i = V_i - \xi$. Conditional on $\mathfrak{F}$, $H(\omega)$ becomes a linear function of the random variable ξ, thus its eigenvalues have the form

$$E_i(\omega) = \lambda_i(\omega) + \xi(\omega),$$

with $\mathfrak{F}$-measurable quantities $\lambda_i(\omega)$. (Naturally, the associated eigenvectors can be chosen constant with respect to ξ.)

Now it is clear that $H(\omega)$ satisfies the following estimate:

$$\begin{aligned}\mathbb{P}\{\,\mathrm{dist}\,(\sigma(H(\omega)),E)\le s\,\} &= \mathbb{E}\,[\,\mathbb{P}\{\,\mathrm{dist}\,(\sigma(H(\omega)),E)\le s\,|\,\mathfrak{F}\,\}\,]\\ &\le N\,\max_i \mathbb{E}\,[\,\mathbb{P}\{\,|\xi-\lambda_i|\le s\,|\,\mathfrak{F}\,\}\,]\\ &\le N\,\sup_{\lambda\in\mathbb{R}} \mathbb{E}\,[\,\mathbb{P}\{\,|\xi-\lambda|\le s\,|\,\mathfrak{F}\,\}\,]\,.\end{aligned}$$

Let $\mathcal{S}$ be the event

$$\mathcal{S} := \{\nu_\xi(s) > 2Ns^{1/2}\}.$$

Next, for any fixed $\lambda\in\mathbb{R}$ and $\delta>0$, we have

$$\begin{aligned}&\mathbb{E}\,[\,\mathbb{P}\{\,|\xi-\lambda|\le s\,|\,\mathfrak{F}\,\}\,]\\ &\le \mathbb{E}\,[\,\mathbf{1}_{\mathcal{S}}\,\mathbb{P}\{\,|\xi-\lambda|\le s\,|\,\mathfrak{F}\,\}\,] + \mathbb{E}\,[\,\mathbf{1}_{\mathcal{S}^c}\,\mathbb{P}\{\,|\xi-\lambda|\le s\,|\,\mathfrak{F}\,\}\,]\\ &\quad\le 2Ns^{1/2} + (2s)^{1/2} = (2N+\sqrt{2})\,s^{1/2}.\end{aligned}$$

Finally, we obtain

$$\mathbb{P}\{\,\mathrm{dist}\,(\sigma(H(\omega)),E)\le s\,\} < (2N+\sqrt{2})N\sqrt{2}s^{1/2} < 5N^2s^{1/2}, \tag{4.1}$$

with $N\ge 2$.

A direct inspection shows that for a Gaussian IID sample $V_1(\omega),\dots,V_N(\omega)$ one would obtain, in the same (indeed, even slightly simpler) way the estimate

$$\mathbb{P}\{\,\mathrm{dist}\,(\sigma(H(\omega)),E)\le s\,\} < \frac{N^{3/2}}{\sqrt{2\pi}}\cdot 2s. \tag{4.2}$$

Certainly, neither of the EVC bounds (4.1) and (4.2) can compete with the optimal Wegner bound [3], available in both cases. However, observe that the above estimates are completely elementary from the functional-analytical point of view, and all technical details are hidden in the probabilistic analysis performed in Section 3. Still, the true value of the main result on the uniform distributions is revealed in the application to the multi-particle EVC bounds, discussed in Refs. [1, 2], where the non-optimal dependence upon N and s (having no visible impact on the final qualitative result) is a small price to pay for the optimal decay bound of the eigenfunctions of the multi-particle Anderson-type Hamiltonians.

References

[1] V. Chulaevsky, *A remark on charge transfer processes in multi-particle systems*, 2010, `arXiv:math-ph/1005.3387`. ↑1, 5, 6

[2] V. Chulaevsky, *On resonances in disordered multi-particle systems*, C.R. Acad. Sci. Paris, Ser. I, **350** (2011), 81–85. ↑5, 6

[3] F. Wegner, *Bounds on the density of states in disordered systems*, Z. Phys. B. Condensed Matter **44** (1981), 9–15. ↑6

Département de Mathématiques, Université de Reims, Moulin de la Housse, B.P. 1039, 51687 Reims Cedex 2, France, E-mail: victor.tchoulaevski@univ-reims.fr